Annotation of Events and Temporal Expressions in French Texts

André Bittar

Annotation of Events and Temporal Expressions in French Texts

André Bittar

Université Paris 7 — ALPAGE

Abstract

An important part of natural language text comprehension is the understanding of temporal information. TimeML is a specification language designed for the markup of temporal information in natural language texts, with an original focus on the processing of English texts.

This article details a pipeline of automatic annotation modules developed for the annotation of temporal information in French texts. We present the results of a preliminary evaluation which yields results comparable to existing systems for English and French. The development of these modules is part of a broader project for the creation of a set of resources for the temporal annotation of French texts according to the TimeML standard.

1 Introduction

The importance of temporal information in natural language texts has received increasing attention in computational linguistics research over the past few years (Mani et al. 2005). Indeed, temporal structure is an essential component of a narrative; events which are described are situated with respect to one another and their participants may evolve over time. Newspaper articles, for example, describe the world around us, mentioning events and states of affairs along with the people involved in them. Furthermore, the order in which events are mentioned in a text does not necessarily represent the chronological order in which they occur. An article will usually mention the most important events first, for example, then continue with descriptions of less important ones which occurred before them. Thus, temporal structure is apparent not only within the bounds of the sentence, but on the level of discourse. This represents a particular challenge for automatic processing as determining the chronological order of events requires analysis of many global factors, including the temporal localization of other events. Furthermore, it is often necessary to use information on many different levels of analysis: morphology, syntax, semantics and pragmatics. Grasping the temporal structure of a narrative is necessary in order to understand a text completely, but is far from being a trivial task.

In the late 1990s, the MUC (Grishman and Sundheim 1996) and, later, ACE (Doddington et al. 2004) information extraction campaigns included a task centered around the recognition and extraction of events. The events in question belonged to a predetermined set of domain-specific relations, such as rocket launchings or company mergers. The temporal coordinates of each event, such as the time and date of occurrence, were to be recorded. These information extraction tasks focused solely on a given domain and the processing of unstructured text was not envisaged in the framework of MUC or ACE.

More recently, the TimeML project (see http://www.timeml.org) has broached the subject of domain-independent processing of temporal information in texts. The TimeML annotation language (Pustejovsky et al. 2003) was developed to render explicit the information concerning events, temporal expressions and the relations which exist between them. A number of resources have been developed for the annotation of English texts and work has also been carried out on other languages, such as Italian, Korean and Chinese. Efforts have been made for the development of resources to process French texts according to the TimeML schema (Bittar 2008, Parent et al. 2008), although they remain scattered. This article details the ongoing development of a coherent set of resources for the annotation of French texts.
Section 2 presents TimeML in further detail, section 3 describes existing resources for French. Section 4 presents modules we are developing for the automatic processing of French texts, namely for the annotation of temporal expressions and the detection and classification of events. The results of a preliminary evaluation are also presented. Section 5 briefly describes other resources under development - an annotation guide and a reference corpus.

2 TimeML

TimeML (Pustejovsky et al. 2003) is a specification language for the annotation and normalization of temporal information in natural language texts. It was principally developed in view of the amelioration of question answering systems, allowing them to process questions of a temporal nature concerning mentioned entities and events. The annotation scheme allows for the following functionalities: the annotation of events, the tagging of temporal expressions and the normalization of their values, as well as the annotation of temporal, aspectual and subordinating relations which may exist among them. The linguistic markers of these relations can also be marked up.
TimeML adopts a broad conception of events, including the traditional conception of events and certain states, corresponding to the notion of eventuality (Bach 1986). These are marked up with the `<EVENT>` tag. This tag allows for the classification of an annotated event according to a predefined ontology. The TimeML event ontology is made up of 7 different classes: `ASPECTUAL` (verbs such as *start, continue, finish...*), `I_ACTION` (*deny, allow, hinder...*), `I_STATE` (*want, believe, like...*), `PERCEPTION` (*hear, see, watch...*), `REPORTING` (*say, quote, state...*), `STATE` (*to be + ADJ*) and `OCCURRENCE` (the events which do not belong to any of the other classes). The `<EVENT>` tag contains attributes to capture grammatical information such as part of speech, verb tense and mood, the presence of a negation, as well as the aspectual properties of a given event.
Markable events can be of several different grammatical categories - verbs, such as *run*, *develop* or *orchestrate*; deverbal nouns, such as *development* or *orchestration*; non-deverbal event nouns, like *war* and *funeral*; adjectives or prepositional phrases introduced as complement of a copula verb, which typically denote states.

The annotation guidelines specify that it is the lexical head of an event-denoting constituent that is to be marked up with this tag. In the examples in 1 below, the lexical item to annotate with the event tag is underlined.

(1) a. *John has <u>eaten</u> four kilos of spaghetti.*
b. *The slow development of Internet technology in New Zealand is frustrating.*
c. *Physicist Stephen Hawking was very <u>ill</u> in hospital last night.*

Temporal expressions are marked up with the `<TIMEX3>` tag. Four main types of temporal expressions can be annotated: dates (type `DATE`, *15 January, 23.02.2008*), times (type `TIME`, *2 o'clock, 15h20*), durations (type `DURATION`, *5 days, two years*) and sets (type `SET`, *every day, each year*). TimeML also provides for reasoning with underspecified deictic temporal expressions such as *next Monday* and *last year*, whose values are resolved with respect to a temporal reference point, for example the publication date of the document.

Events and temporal expressions are related to each other by three types of link tags: temporal links (`TLINK`), aspectual links (`ALINK`) and subordinating links (`SLINK`). The first captures temporal relations, such as precedence, simultaneity or inclusion, between two entities (`EVENT-EVENT`, `TIMEX3-TIMEX3`, or `EVENT-TIMEX3`). The second captures the phases in the unfolding of an event, for example the beginning, the middle or the end. The third is essential for reasoning which depends on the truth conditions or the certainty of event denoting propositions, such as in the case of certain modal verbs, for example. Functional words which explicitly mark one of these relations are annotated with the `SIGNAL` tag. These are most often prepositions like *before, after* or *during*.

A set of resources, including automatic and manual annotation tools and several reference corpora, available on the project website, have been developed around the TimeML language. Evita (Saurí et al. 2005) is an application for automatically recognizing and annotating events in texts, based primarily on symbolic methods and linguistic data, although with some integration of statistical data. Evita takes as input a text which has undergone part-of-speech tagging, morphological analysis and chunking. The authors report precision of 74.03% and recall of 87.31% for an overall F-score of 80.12% for the task of event identification across all the relevant grammatical categories. These results are comparable to the inter-annotator agreement scores given by the authors for this task: 64% for identification of nominal events and 80% for event verbs. The GUTime application (Mani and Wilson 2000) annotates temporal expressions according to the TimeML schema and normalizes their values. The system achieves F-scores of 85% and 82% for identification and normalization of temporal expressions, respectively. Three separate modules Slinket, GutenLink and S2T identifiy and annotate subordinating and temporal relations. Information on these modules is available on the TimeML website.

Three different hand-annotated corpora have been produced for English: TimeBank 1.1, TimeBank 1.2 and the Aquaint TimeML corpus. Figures for these corpora are presented in Table 1.

	TimeBank 1.1	TimeBank 1.2	Aquaint TimeML	Combined corpora
Tokens	64 887	61 000	31 000	156 887
`TIMEX3`	1263	1414	605	3282
`EVENT`	3171	7935	4432	15 538
`SIGNAL`	1497	688	268	2453
`ALINK`	294	265	365	924
`SLINK`	2743	2932	3932	9607
`TLINK`	5957	6418	5365	17 740

Table 1: Figures for TimeML annotated corpora for English

Over recent years, TimeML has been adopted by the International Organization for Standardization and efforts are under way to provide an ISO standard for the annotation of temporal information, called ISO-TimeML. Annotation guidelines have been written for various languages to deal with language-specific phenomena. At present, the languages other than English which have received attention are Italian, Korean and Chinese (ISO 2008).

3 Existing Resources for French

Resources for the annotation of French texts in TimeML are relatively few and far between. Although some efforts have been made, the resulting work remains scattered. (Muller and Tannier 2004, Muller and Reymonet 2005) focus on the annotation of temporal relations between events, proposing algorithms and a methodology for evaluation.

As far as annotation systems are concerned, (Parent et al. 2008) provide the description and evaluation of a system for the TimeML annotation of events and temporal expressions in French texts. The processing of temporal expressions is carried out on a text having undergone a part-of-speech tagging, morphological analysis and shallow syntactic analysis. The system functions by application of a cascade of 90 rules applied over 20 levels. The syntactic information is only used when calculating the value of ambiguous expressions (for example *14 heures* - meaning either *2 p.m.* or *14 hours*), or if the context is required for interpretation. Little detail is given as to how the normalization process is carried out.

Contrary to the Evita system developed for English, the event detection module relies on a full dependency parse as input for the event recognition task. The authors claim an advantage over chunker-based approaches with respect to the annotation of markable adjectives due to the fact that the dependency relation between copula verb and predicative adjective is provided by the parsing.

The authors provide evaluation results according to grammatical category over a development corpus, made up of 35 biographical texts and 22 sports articles, and an evaluation ("unseen") corpus, consisting of an unspecified number of news articles from the website of the École Polytechnique de Montréal. Figures for the two corpora, in Table 1, show that they are much smaller than those available for

English. The evaluation results, by grammatical category and global figures, are given in Table 3.

	Development corpus	Evaluation corpus
Documents	35	Unspecified
Sentences	400	90
Words	9 234	2 673
`EVENT`	840	210
`TIMEX3`	419	87
`SIGNAL`	210	47

Table 2: Figures for corpora used

	Development corpus			Evaluation corpus		
Category	Precision	Recall	F-score	Precision	Recall	F-score
Noun	61.5	40.0	48.4	54.7	53.7	54.2
Verb	94.1	97.3	95.7	65.6	90.9	76.2
Adjective	66.7	77.8	71.8	N/A	N/A	N/A
Global	86.8	80.6	83.5	62.5	77.7	69.3

Table 3: Evaluation results according to corpora

The system performs best on the annotation of event verbs and encounters the most difficulties in the annotation of event nominals, for which the TimeML annotation guidelines are fairly minimal. Adjectives are relatively well processed over the development corpus, but no adjectives were annotated by the human annotator in the evaluation corpus, so no results were calculated.

As for the recognition and annotation of temporal expressions, this application achieves a precision of 83% and a recall of 79% for an F-score of 81% over an evaluation corpus containing 544 human-annotated temporal expressions and an F-score of 50% for the normalization of values. These figures are comparable to those cited for the GUTime application for English (see section 2).

4 Annotation Modules

In this section, we describe an annotation system, similar to that of (Parent et al. 2008) described above, although based on a rich cascade of finite state transducers and the output of an "augmented" chunker (a shallow syntactic analysis) as opposed to a full dependency parse. The system is made up principally of two modules, the first tagging temporal expressions (section 4.1), the second identifying and annotating event expressions (section 4.2).

4.1 Temporal Expression Tagger

This module carries out the tagging and normalization of temporal expressions, with the `<TIMEX3>` tag and certain relation markers with the `<SIGNAL>` tag. `<SIGNAL>`s marked up with this module are those which introduce a temporal expression, such as *pendant deux semaines* (*for two weeks*) or *après lundi* (*after Monday*). The module is based on the Time-French package of graphs for matching temporal expressions in French, developed with the Unitex[1] corpus processor by Maurice Gross (Gross 2002). The graphs in this large-coverage grammar recognize patterns of dates, times, duration and frequency. However, they do not output the type of the recognized expressions nor the other information required by TimeML. The graphs had to be completely reorganized in order to represent the temporal expressions recognized within the TimeML schema. For example, the graphs of Time-French did not represent the same classification as TimeML. For example expressions of duration were sometimes matched by the same graph as times. The graphs therefore had to be separated into sets corresponding to the types of temporal expressions recognized by TimeML. Moreover, numerous temporal expressions, such as *à l'heure convenue* (*at the agreed time*), *en toute éventualité* (*in any eventuality*) and *l'ère paléozoïque* (*the Paleozoic era*) were removed as they do not fit within the framework of calculable expressions targeted by TimeML. The appropriately typed `<TIMEX3>` and `<SIGNAL>` tags were added to the graphs. This equates to turning finite state automata into transducers.

A graph matching expressions **not** to be marked up was also created. This graph tagged with the label `<GARBAGE>` expressions such as phone numbers (*06.33.08.12.74*), which could otherwise match numerical dates. The ambiguous word *été* (*been/summer*), when preceded by an adverb or the auxiliary verb *avoir* is tagged as `<GARBAGE>`, as it has its verb rather than noun reading in this context. Age expressions such as *âgée? (de|d') NUM ans?*, *l'âge (de|d') NUM ans?* and *NUM ans? d'âge*, where NUM represents an alphabetical or numerical number expression, were also tagged in this way in order to avoid tagging them as duration expressions. Other expressions tagged as `<GARBAGE>` include the common expression *les 35 heures* (*the French 35 hour week*), geographical references containing the word *Midi* (the south of France or *midday*), such as *Midi-Pyrénées* and *Midi-Atlantique*, and names of streets containing a date, such as *la place du 13 Mai*, etc. Annotations on expressions tagged as `<GARBAGE>` in the text are subsequently removed by the script which performs the normalization of values.

The transducers are applied to raw text to output a text marked up with the `<TIMEX3>` tag. The normalization script, written in Perl, calculates the standard values of temporal expressions, including underspecified deictic expressions. The reference date from which to calculate these values is able to be entered by the user. The script consists of a set of substitution functions for each type of temporal expression tagged by the transducers. Each function converts the content of the expression into a TimeML standard value and inserts it in the

---

[1] Unitex is a graphical corpus processing program, available for download under GNU General Public Licence at http://www-igm.univ-mlv.fr/~unitex/

value attribute of each <TIMEX3> tag. This package is available for download at http://www.linguist.univ-paris-diderot.fr/~abittar.
This approach differs from that of (Parent et al. 2008) in that it relies solely on lexical processing. Evaluation results, presented below, indicate that lexical processing yielded similar, and in some cases better, results to the approach based on more complex computation.

Evaluation of the temporal expression tagger was carried out on a subset of the corpus used to evaluate the similar module described in section 3[2]. Our corpus includes all documents used in the training and test sets described by (Parent et al. 2008), adapted from the one used in (Baldwin 2002). It consists of 45 news articles from the Agence France Press, with a total of 592 human-annotated <TIMEX3> tags. These texts were not used in development and thus contain completely "unseen" material. Figures for the evaluation are given in Table 4. The column labeled "Loose match" represents the number of approximate matches which cover an incomplete span of the expression, for example *un mois* (*one month*) instead of *un mois et demi* (*a month and a half*) or *vingt-quatre heures* (*twenty-four hours*) instead of *les dernières vingt-quatre heures* (*the past twenty-four hours*). The column "Strict match" is for exact matches on the span of the expression. The "Correct value" column represents the correctly normalized values for the temporal expressions detected, calculated over the strict matches.

	Human	Found	Loose match	Strict match	Correct value
Number	592	575	508	484	317
Precision	-	-	85.8	84.2	55.0
Recall	-	-	88.4	81.8	44.9
F-score	-	-	87.1	83.0	49.4

Table 4: Evalutation results for the temporal expression tagger

The evaluation of our module indicates a performance much in line with that of the system described in (Parent et al. 2008). Performance is slightly lower on loose matches (F-score 87.1 versus 91.0), but we achieve better results on strict matches (F-score 83.0 versus 81.0). This could be explained by the fact that we did not develop our grammar on the same type of source text, but shows that the grammar has a good coverage of the variants of each type of expression.
Sources of noise include age values tagged as durations (2a) (11 errors) and numerical values taken to be years (2b) (8 errors), while silence occurs mostly on coordinated date expressions or sequences (2c) (11 errors) or expressions taking a "vague" normalized value (2d) and (2e) (15 errors).

(2) a. *M. Dupont, 58 ans*
Mr. Dupont, 58 years old

[2]We have not evaluated annotation of the SIGNAL tag as this was not carried out for the module with which we compare our results.

b. *L'astéroïde 2001 UU92*
Asteroid 2001 UU92
c. *Les 4, 5 et 6 février.*
The 4th, 5th and 6th of February.
d. *Dans le passé*
In the past
e. *À l'avenir*
In the future

Results for the normalization of values for temporal expressions are practically identical to the other system for French. A large part of the errors produced by our system (97 out of 167) are due to the fact that our normalization script does not yet fully deal with underspecified weekday expressions, such as *jeudi soir* (*Thursday evening*). In the hand-annotated corpus these expressions are fully resolved, with year, month and day values specified, e.g. `2002-01-15TEV`, whereas we provide a correct, but not completely resolved value, which specifies the day of the week, e.g. `2002-WXX-4TEV`. Excluding this difference in processing boosts precision to 73.6 and recall to 60.1 (F-score 66.85) for the normalization of values. We are currently working on fully normalizing these values.
Another type of error (7 errors) occurs in incorrectly resolved year values for month expressions such as *novembre* (*November*), which resolve to the year of the article's publication date instead of the preceding or following year. For example, in (3a), the month *décembre* refers to the month of December in the past, this being indicated by the past tense of the verb. As grammatical tense information is not available to the normalization script, this fact is not taken into account. 12 errors come from the fact that the normalization script does not yet deal with references to parts of the business year, such as in (3b). A further 22 errors occurred for expressions with a "vague" reference, such as *aujourd'hui* (*today*) in (3c), which refers to the present day in general rather than the specific day's date. Our system resolves these with specific date values, rather than using the values `PAST_REF` for vague references to the past, `PRESENT_REF` (for vague references to the present) or `FUTURE_REF` (for vague references to the future). Thus, *aujourd'hui* is normalized to the specific value `2002-03-21` (the date of publication of the article) instead of `PRESENT_REF`.

(3) a. *Le site a dégagé un réel bénéfice en décembre.*
The site made a real profit in December.
b. *Le chiffre d'affaires a augmenté de 8% au deuxième semestre 2002.*
Turnover increased by 8% in the second semester of 2002.
c. *Aujourd'hui, un tiers des ventes s'effectue sur Internet.*
Today, a third of sales are made over the Internet.

The expression *le 11 septembre* (*September 11* - i.e. *2001*) has 4 occurrences which are incorrectly resolved to the year of the publication date of the article. Correctly resolving these expressions would rely on contextual cues such as the head noun in *les attentats du 11 septembre* (*the September 11th attacks*).

Although these last two types of errors are not very frequent, they do show the possible benefits of a broader contextual analysis for the processing of temporal expressions.

4.2 Event Tagger

This module tags event expressions with the <EVENT> tag and classifies the events according to the ontology defined for TimeML. It also detects negative polarity contexts, as well as any aspectual or modal properties of certain verbal constructions. In annotating event expressions, the lexical head of the event expression is annotated, as the examples in 4 show (the attributes of the <EVENT> tags are omitted):

(4) a. Le chat a <EVENT>mangé</EVENT> la souris.
The cat ate the mouse.
b. La violente <EVENT>destruction</EVENT> de la ville.
The violent destruction of the city.
c. Jean était <EVENT>malade</EVENT> pendant deux jours.
Jean was sick for two days.

The event tagger also marks up certain relation markers with the <SIGNAL> tag, namely those which directly introduce an event expression, such as in *lors de l'évacuation* (*during the evacuation*) or *après avoir mangé* (*after having eaten*).
A certain amount of pre-processing is necessary before the application of this module. It's input is a text having undergone segmentation into sentences, tokenization, part-of-speech tagging, an inflectional morphological analysis and shallow syntactic analysis. This pre-processing is carried out by Macaon, a modular processing pipeline for French[3].
The event tagger consists of several levels of processing. The first level is a layer of lexical processing, basically a lexical lookup for nouns and verb classes. The second layer is for contextual processing consisting in the application of heuristics for detecting and eliminating event candidates and classifying them.
This module relies on certain lexical resources. For the detection of event nominals, a lexicon containing nouns with at least one event interpretation is used. Many of the entries in this lexicon are ambiguous as they may also have a non-event interpretation. For example, *présentation* (*presentation*), *description* and *repas* (*meal*) have object or information interpretations as well as event readings. The necessity for disambiguation of nominals is therefore apparent. Several sources contributed to the creation of this lexicon. The first source of potential event nominals is the VerbAction lexicon (Hathout et al. 2002) which contains 9 393 verb-deverbal noun pairs, giving a total of 9 200 unique nominal lemmas (used in our lexicon). To complement these entries, we semi-automatically extracted, via search engine queries, the elements appearing in patterns such as *X a eu lieu* (*X took place*), *X s'est produit* (*X happened*), *lors du/de la/des X* (*during*

[3]Macaon is freely available for download at http://pageperso.lif.univ-mrs.fr/~alexis.nasr/macaon/.

the X), *le X de...par...* (*the X of ...by...*), where X is likely to be an event nominal. An initial application of this method yielded 769 unique noun lemmas which were not in VerbAction, mostly rare or non-deverbal nouns, such as *anticoagulothérapie* (*anticoagulation therapy*) and *anniversaire* (*birthday*). This method is to be regularly reapplied in order to progressively increase lexical coverage. The noun lexicon is of comparable size to that used in the Evita application for English (13 495 entries), described in section 2.

We created by hand a verb lexicon which is used to perform classification of verbal events. It contains 200 lemmas for verbs in 6 of the 7 TimeML event classes. As the class `OCCURRENCE` is the default class, it has no entries in the lexicon. Verbs were initially added to the lexicon by translating those proposed in the TimeML classifcation for English. The list of verbs was enriched by querying the dictionary of synonyms at the Université de Caen (http://www.crisco.unicaen.fr/cgi-bin/cherches.cgi). The lexicon is small for the time being and will need to be increased to ensure better coverage for classification, although the vast majority of verbs belong only to the default class and do not need to appear in the lexicon. Like the noun lexicon, the lexicon of verbs contains ambiguities as certain verbs may belong to different classes or may not have an event reading in certain contexts. For example, the verb ***expliquer*** (***to explain***) belongs to the class `REPORTING` when it introduces either a complementizer phrase in *que* (*that*) headed by an event (5a) or direct speech (5b). This is the class attributed by the lexicon. However, when it has a human subject and an event in object position (5c), it must be annotated with the class `I_ACTION`. Finally, if this verb has events in both subject and object position (5d), it is to be annotated with the class `CAUSE`.

(5) a. Ballmer a expliqué que Microsoft ne commettrait pas cette erreur.
Balmer explained that Microsoft would not make this mistake.
b. “La situation risquerait de s’aggraver”, a expliqué le porte-parole.
“The situation may become more serious”, explained the spokesperson.
c. L’ex-capitaine a expliqué le renouvellement de l’équipe.
The ex-captain explained the renewal of the team.
d. Le réchauffement climatique explique la fonte des calottes glaciaires.
Global warming explains the melting of the ice caps.

The system is thus confronted with the non-trivial problem of word sense disambiguation to identify the correct readings of nouns and verbs in the text. Initially, we tackle this problem for verbs with a number of heuristics, applied to local chunk context, for each of the TimeML verb classes in the lexicon. A total of 16 heuristics are used for choosing candidates for markup with the `<EVENT>` tag and 30 heuristics for classifying the events (`class` attribute) and determining values for the `aspect`, `modality` and `polarity` attributes. For example, in the case of the verb *expliquer* given above, the heuristics include a search for the complementizer *que* in the chunk after the verb, a search for cited text (quotation marks) to the left of the verb and a search for an event nominal chunk directly to the left of the verb chunk (approximation of subject position).

Further heuristics are used to eliminate verbs and nouns which do not have an event reading. For example, event nominal chunks which do not have a determiner, such as in *prisonier de guerre* (*prisoner of war*), are not considered as candidates as they do not denote event instances, but rather event types, and cannot be attributed a specific temporal localisation. A set of heuristics is used to detect predicative adjectives, like in *Jean était malade* (*Jean was sick*), which are potential candidates for markup with the `<EVENT>` tag. For example, if the preceding verb is a copula, the adjective is flagged as a markable.

The corpus used to evaluate the temporal expression tagger was not annotated for events. To evaluate our event tagger we used a corpus of 30 hand-annotated news articles from the newspaper *Le Monde*. The corpus was split into a development set of 20 documents (11 224 tokens, 1 187 `EVENT` tags) and a test set of 10 documents (5 916 tokens, 583 `EVENT` tags). Overall, the corpus contains 1 205 verbal, 471 nominal, 62 adjectival and 18 prepositional phrase `EVENT` tags.

	Development corpus			Evaluation corpus		
Category	Precision	Recall	F-score	Precision	Recall	F-score
Noun	50.2	94.5	72.4	54.0	95.1	74.5
Verb	87.7	92.3	90.0	86.5	91.1	88.8
Adjective	60.0	72.4	66.2	46.0	82.1	64.1

Table 5: Evalutation results for the event tagger

The results shown in Table 5 are fairly homogenous over both the development and test sets, which were taken from the same news source. This is reassuring as it indicates that the heuristics used may be general enough to be effective across different types of texts - although we will carry out further tests on different types of corpora to confirm this. The detection of event verbs performs slightly lower than that of the other system for French, although the evaluations were carried out independently and on different corpora.

There is a stark contrast in results for the annotation of event nominals. Our system makes a vast improvement on the performance of the other system described in this paper (an F-score of 74.5 versus 54.2 over the respective test sets). The large-coverage lexicon of event nominals allows for a good recall, although precision remains relatively low as more disambiguation techniques are required to further filter out nominals with non-event readings.

Performance on adjectival `EVENT`s is lower than the other system, although not as bad as might have been expected. The difference is certainly due to the depth of syntactic analysis on which the systems are based - our system relies on simple chunk-level heuristics whereas the other benefits from a full dependency parse which links predicative adjectives to their governor. The processing of adjectives and nouns are the areas which require the most attention, although, proportionately, markable nouns are vastly more common than markable adjectives.

Although these results give an idea of the performance, for a more meaningful comparative evaluation, the systems will have to be tested on the same corpus.

The need for a validated reference corpus is apparent. Furthermore, to complete the evaluation the other aspects of the task at hand, namely the classification of events and the detection of aspectual, modal and negative polarity properties, will need to be carried out. These are issues we seek to address in future.

5 Annotation Guide and Reference Corpus

As not all languages express notions of time in the same way, certain adaptations to the TimeML guidelines are necessary in order to account for language-specific phenomena. We have devised a set of guidelines for French, which have been validated by other researchers within the French-speaking computational linguistics community. Without going into detail, so as not to stray from the scope of this article, the guidelines cover, among other phenomena, the treatement of modality, aspectual expressions and causative constructions, as well as the annotation of event nominals, which are lacking in the general TimeML annotation guide.
With our guidelines in mind, a Gold Standard reference corpus was produced using the annotation tool Callisto[4] as a basis for evaluation of the event tagger module (details are given in section 4.2). Due to the extremely time-consuming nature of the annotation, the corpus is relatively small. Furthermore, due to copyright restrictions, it is not freely available. This is a problem common to all the corpora currently annotated for French as far as we know. A viable solution does exist, however. The Centre national des ressources textuelles et lexicales (http://www.cnrtl.fr) provides for free download 3 years of issues of the newspaper *L'Est Républicain* (roughly 125 million words). We are currently in the process of annotating a small subset of this corpus. This corpus consists of one day's issue of the newspaper, approximately 350 000 tokens, just over twice the size of the combined corpora for English. Complete manual annotation of such a volume of text is impracticable and the methodology we adopt to carry out the task is to annotate automatically using the modules described in section 4 and perform a manual correction of the output. The entire corpus will be annotated for `EVENT`s, `TIMEX3`s and `SIGNAL`s. As the automation of the annotation of relations is much more complicated than the tasks of event recognition and detection of temporal expressions, a small subset of the whole corpus will be annotated with the relation `ALINK`, `SLINK` and `TLINK` tags. The resulting corpus will be of adequate size and quality to provide a reasonable starting point from which to train machine learning algorithms.

6 Conclusion

This article has focused on presenting two modules for the automatic processing of temporal information in French texts according to the TimeML markup language. We have presented an evaluation of a module for the annotation and normalization of temporal expressions as well as a module for the detection and classification of events, comparing, where possible, with an existing system. Results are similar for

[4] Available at http://callisto.mitre.org

the detection and normalization of temporal expressions. For event detection, our system performs slightly lower for verbs and adjectives, but significantly higher for event nominals. Although figures from the evaluation give a rough indication of performance across systems, a validated reference corpus for French is yet to be developed in order to give more meaningful comparisons. This is an issue we are currently addressing through the creation of an annotation guide and a reference corpus for French. Our aim is to eventually provide a coherent set of resources which will be available to the scientific community.

References

Bach, Emmon (1986), The Algebra of Events, *Linguistics and Philosophy* **9** (1), pp. 5–16.

Baldwin, Jennifer (2002), *Learning Temporal Annotation of French News*, Master's thesis, Georgetown University, Washington DC, USA.

Bittar, André (2008), Annotation des informations temporelles dans des textes en français, *Actes de RECITAL 2008*, Avignon, France.

Doddington, George, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel (2004), The Automatic Content Extraction (ACE) Program Tasks, Data, and Evaluation, *Proceedings of LREC, 2004*, Lisbon, Portugal.

Grishman, Ralph and Beth Sundheim (1996), Message Understanding Conference - 6: A Brief History, *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Vol. 1, Copenhagen, Denmark, p. 466471.

Gross, Maurice (2002), Les déterminants numéraux, un exemple: les dates horaires, *Langages* (145), pp. 21–38, Larousse, Paris, France.

Hathout, Nabil, Fiammetta Namer, and Georgette Dal (2002), An Experimental Constructional Database: The MorTAL Project, *in* Boucher, Paul, editor, *Many Morphologies*, Cascadilla, Somerville, Mass., USA, pp. 178–209.

ISO, S.W.G. (2008), *ISO DIS 24617-1: 2008 Language resource management - Semantic annotation framework - Part 1: Time and Events*, International Organization for Standardization, ISO Central Secretariat, Geneva, Switzerland.

Mani, Inderjeet and George Wilson (2000), Processing of News, *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL2000)*, Association for Computational Linguistics, Hong Kong, pp. 69–76.

Mani, Inderjeet, James Pustejovsky, and Robert Gaizauskas (eds.), editors (2005), *The Language of Time, A Reader*, Oxford University Press.

Muller, Philippe and Axel Reymonet (2005), Using Inference for Evaluating Models of Temporal Discourse, *Proceedings of the 12th International Symposium on Temporal Representation and Reasoning*, IEEE Computer Society Press, pp. 11–19.

Muller, Philippe and Xavier Tannier (2004), Annotating and Measuring Temporal Relations in Texts, *Proceedings of Coling 2004*, Vol. 1, Association for Computational Linguistics, Geneva, Switzerland, pp. 50–56.

Parent, Gabriel, Michel Gagnon, and Philippe Muller (2008), Annotation d'expressions temporelles et d'événements en français, *Actes de TALN 2008*, Avignon, France.

Pustejovsky, James, José Casta no, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz (2003), TimeML: Robust Specification of Event and Temporal Expressions in Text, *Proceedings of IWCS-5, Fifth International Workshop on Computational Semantics*.

Saurí, Roser, Robert Knippen, Marc Verhagen, and James Pustejovsky (2005), Evita: A Robust Event Recognizer for QA Systems, *Proceedings of HLT/EMNLP 2005*, pp. 700–707.